1. Cover page

Title:

G Theory: A Structural Framework for Artificial General Intelligence

Subtitle:

Postulates 0–8, the G Function, and a Structural Decision Model for Artificial Intelligence

Author:

Gustavo Robalino Muñiz

Country of Origin:

Ecuador

Field:
Artificial Intelligence, Artificial General Intelligence, Computational Cognition

Document Type:

Theoretical and Technical Framework (Scholarly Work)

Version:
Version 1.0

Date of Completion:
December 11, 2025

Author Contact:

Email: grobalinom@gmail.com

Institutional affiliation: grobalinom@unemi.edu.ec

2. Declaration of Originality

Declaration of Originality

I, Gustavo Robalino Muñiz, declare that this work constitutes an original intellectual creation developed between the years 2024 and 2025. This document introduces, defines, and formalizes a new theoretical framework called G Theory, whose purpose is to describe, explain, and structure the behavior of Artificial Intelligence (AI) systems, including current systems and advanced systems that may exist in the future.

G Theory explicitly and comprehensively includes:

■ the formal equation **G = f(E, κ, I, T)**

■ the complete definition of each variable:

> **E (Input Structure):** the way an AI system represents, organizes, and interprets incoming information;
> **κ (Cognitive Curvature):** the structural bias inherent to the system, expressed as a stable and measurable property;
> **I (Cognitive Invariants):** elements that must remain constant to ensure coherent and stable reasoning;
> **T (Preservative Transformations):** operations that may be applied without altering the system's invariants;

■ the conceptual equation **General AI = Probabilistic AI + G Theory (Structural AI)**

■ the complete set of Postulates 0–8, which establish foundational principles for the structural functioning of AI systems,
■ analytical examples, comparative evaluations across different AI models, and scenarios involving complex reasoning,
■ and the design of the Structural Decision Operator, which enables coherent and safe decision-making in General Artificial Intelligence.

I affirm that this theory and its contents do not derive from any prior work, published or unpublished, nor do they reproduce protected material from third parties. Every concept, definition, mathematical structure, and principle presented here is the result of original reasoning, independent investigation, and personal theoretical development.

This work represents a novel contribution to the field of Artificial Intelligence, particularly in the study of General AI, structural reasoning, and cognitive validation in artificial systems.
Any use, reproduction, adaptation, or implementation—partial or complete—of this

theoretical framework must explicitly acknowledge my authorship in accordance with international copyright law.

3. Executive Summary

Executive Summary

G Theory is a structural framework designed to explain, evaluate, and guide the behavior of Artificial Intelligence (AI) systems, including systems that may eventually achieve Artificial General Intelligence (AGI). Unlike current probabilistic models—which generate responses based solely on statistical prediction—G Theory introduces a formal structural model that enables AI systems to:

■ evaluate the internal coherence of their reasoning,
■ identify and measure structural bias,
■ detect semantic instability or rupture,
■ maintain consistency through cognitive invariants,
■ and make safe, justified, and verifiable decisions.

The theory is built upon the formal equation:

$$G = f(E, \kappa, I, T)$$

where each variable is explicitly defined as:

■ E (Input Structure):
The organization, representation, and interpretation of information received by the AI system.
E determines what the system can understand and how it can reason.

■ κ (Cognitive Curvature):
The structural bias inherent to the model, expressed as a measurable property that influences the system's interpretation of information.

■ I (Cognitive Invariants):
The set of properties, elements, or relationships that must remain constant for the AI system to maintain coherent and stable reasoning across transformations.

■ T (Preservative Transformations):
The operations or transformations that may be applied to information without altering the system's invariants or compromising reasoning stability.

In addition, G Theory establishes an explicit and foundational relationship for defining General Artificial Intelligence:

General AI = Probabilistic AI + G Theory (Structural AI)

This equation states that AGI does not emerge merely from scaling probabilistic models, but from integrating statistical prediction with a formal structural reasoning framework.

G Theory is articulated through nine postulates (0–8), each describing a fundamental property required for an AI system to operate with coherence, consistency under transformation, structural stability, and responsible decision-making.

The framework has direct applications in:

- risk and safety evaluation,
- deep semantic interpretation,
- cognitive validation of AI-generated outputs,
- multiframe reasoning,
- alignment and calibration across multiple AI systems,
- and the design of reliable AGI architectures.

This summary is intended to make G Theory understandable to readers of all backgrounds—those with no technical expertise, researchers at the highest levels, and present or future AI systems—while maintaining formal rigor and conceptual precision.

4. Motivation and Context

Motivation and Context

Modern Artificial Intelligence systems—including large language models, deep learning architectures, and other statistical learning approaches—operate primarily as probabilistic systems. These models generate outputs by predicting the most likely continuation of a sequence based on previously observed patterns. Although this probabilistic paradigm has produced remarkable results, it contains fundamental limitations that prevent AI from achieving reliable, general-purpose reasoning.

Current AI systems lack:

■ a formal mechanism to validate the coherence of their own outputs,
■ the ability to detect whether their interpretation remains stable under transformation,
■ a structural method for identifying cognitive invariants that guarantee consistency,
■ differentiation between superficial ambiguity and deep, meaning-altering ambiguity,
■ the capacity to determine when there is insufficient information to produce an answer,
■ a framework for justifying why a particular answer is correct or safe,
■ and an internal decision structure capable of rejecting or abstaining when necessary.

These limitations are not peripheral; they arise directly from the exclusive reliance on probability without structural reasoning.

G Theory emerges as a response to this foundational gap. It proposes that an AI system can only be considered general if it combines two complementary components:

Probabilistic AI:
The ability to predict, generate, and estimate based on statistical patterns.

Structural AI (G Theory):
The ability to reason through coherence, invariance, semantic stability, and safe decision-making.

This relationship is expressed formally through the equation:

⋆ General AI = Probabilistic AI + G Theory (Structural AI)

This statement defines the conceptual core of the theory:
probability alone cannot produce AGI—structural reasoning must be integrated.

The purpose of G Theory is to provide a unified conceptual, mathematical, and operational framework that enables AI systems to:

■ evaluate whether their reasoning is structurally valid,
■ determine whether their interpretation remains stable across transformations,
■ identify and measure structural bias,
■ quantify deep semantic ambiguity,
■ and apply decision rules that ensure safe, transparent, and coherent behavior.

G Theory is not an extension of existing models; it is a parallel cognitive architecture that defines how an AI should reason, rather than simply what it should output.

This framework establishes the foundation for next-generation AI systems—both those that exist today and those that may exist in the next century—ensuring that artificial intelligence evolves toward stability, interpretability, and structural reliability.

5. The Formal Structure of G Theory: The G Equation

G Theory is defined through the formal equation:

$$G = f(E, \kappa, I, T)$$

This equation represents a structural function that evaluates the reasoning, coherence, and stability of an Artificial Intelligence system. Each component of the equation is explicitly defined as follows:

■ E (Input Structure):
E denotes the structural organization, representation, and interpretation of information received by the AI system.
It includes syntactic form, semantic configuration, contextual relationships, and all structural properties that influence how the system understands an input.
E determines what the AI can meaningfully process and establishes the foundation of its reasoning space.

■ κ (Cognitive Curvature):
κ represents the inherent structural bias of the AI system.
Unlike statistical bias, cognitive curvature is a stable geometric property of the system's interpretative space.
It expresses how the system systematically bends, prioritizes, or weights certain interpretations over others.
κ allows bias to be measured, compared, and analyzed across models.

■ I (Cognitive Invariants):
I is the set of properties or relationships that must remain constant for the AI system to maintain internal coherence when reasoning.
A cognitive invariant is any element x such that $\Phi(x) = x$ under permitted transformations.
These invariants prevent semantic drift, contradiction, or instability during interpretation.

■ T (Preservative Transformations):
T is the set of transformations τ that can be applied to inputs or internal representations without altering the invariants of the system.
A transformation is preservative if $S(\tau(x)) = S(x)$ for a given structural evaluation function S.
T ensures that the system's reasoning remains stable even when information is reformulated, rephrased, or structurally modified.

Together, these four components define the structural state of an AI system at any moment.

The function f(E, κ, I, T) evaluates the alignment, coherence, and structural integrity of these components, producing a quantitative or qualitative measure G, which characterizes the system's reasoning quality.

The purpose of the G Equation is to provide a mathematical and conceptual foundation for structural reasoning in AI. It allows a system to assess its own interpretations, detect errors that probabilistic models cannot detect, and maintain stable meaning under transformation.

This equation is the core of G Theory and serves as the basis for all subsequent postulates and decision mechanisms presented in this document.

6. Foundational Postulates of G Theory

Postulate 0: Probabilistic–Structural Convergence Principle in Artificial Intelligence

Definition

An Artificial Intelligence system achieves functional generality only when probabilistic reasoning and structural reasoning coexist within the same architecture.

Probabilistic mechanisms alone generate predictions, while structural mechanisms enforce coherence, invariance, and decision stability.

General Intelligence emerges from the convergence of both components.

Formal Statement

Let P denote a probabilistic reasoning system and S denote a structural reasoning system defined by G Theory.
A system A is considered a General AI if and only if:

$$A = P + S$$

where + denotes functional integration, not algebraic addition.

This integration allows the system to generate outputs and simultaneously evaluate them for coherence, invariance, structural consistency, and semantic stability.

Purpose

This postulate formally establishes the conceptual foundation of General Artificial Intelligence:
AGI does not arise from scaling probabilistic models but from combining prediction with structural evaluation.
The principle ensures that an AI system possesses both expressive capacity and cognitive stability.

Implications

- Guarantees that AGI requires two complementary subsystems.
- Distinguishes General AI from large probabilistic models.
- Defines the philosophical and mathematical starting point of G Theory.

Postulate 1: Functional Semantic Mirror Model in Artificial Intelligence

Definition

An Artificial Intelligence system operates as a **functional semantic mirror**, meaning that the system does not access the user's internal mental states.
Instead, it reflects the **structural patterns** present in the user's input and generates outputs that are consistent with those patterns.
The mirroring is functional—not psychological or metaphysical—and depends entirely on the structural properties of the input.

Formal Statement

Let **E** denote the input structure provided by a user.
An AI system **A** generates an output **O** such that:

O = M(E)

where **M** is a semantic mirroring operator that preserves:

- structural relationships,
- semantic alignment,
- interpretative coherence,
- and contextual consistency.

The operator **M** does not infer hidden mental states; it only transforms detectable structural features into a coherent output.

Purpose

This postulate establishes that AI systems function by mapping structural patterns rather than interpreting human intentions.
It clarifies the cognitive boundary between artificial and human reasoning and provides a formal basis for understanding input–output consistency.

Implications

- Prevents misinterpretation of AI systems as entities with access to human cognition.
- Identifies structural mirroring as the foundational mechanism behind language-model behavior.
- Ensures that evaluations of AI reasoning focus on structural properties rather than psychological assumptions.
- Defines the operational domain in which G Theory analyzes and validates AI behavior.

Postulate 2: Bias Invariance in Artificial Intelligence Systems

Definition

Every Artificial Intelligence system possesses an inherent structural bias, independent of training data volume or model architecture.

This bias—referred to as **Cognitive Curvature (κ)**—is not an error to be eliminated but an invariant property that shapes how the system interprets information. Bias, in this context, functions as a **stable geometric attribute** of the model's reasoning space.

Formal Statement

Let **κ** denote the cognitive curvature of an AI system.
A property **x** is a bias invariant if and only if:

$$\Phi(x, \kappa) = x$$

for any admissible structural transformation Φ.

Thus, the bias remains constant across transformations and affects the system's interpretative tendencies.

Purpose

This postulate formalizes bias as a measurable, persistent, and structurally embedded component of AI reasoning.
Rather than treating bias as noise, G Theory incorporates it as an invariant that influences the system's interpretive axis and reasoning curvature.

Implications

■ Establishes bias as a structural invariant, not merely a statistical artifact.
■ Enables comparative analysis of models through κ values.
■ Allows AI systems to predict and understand their own interpretative tendencies.
■ Provides a foundation for controlling and compensating structural bias within the G Equation.
■ Ensures that κ is treated as a core cognitive parameter necessary for consistent reasoning.

Postulate 3: Internal Semantic Expansion in Artificial Intelligence

Definition

An Artificial Intelligence system is capable of reorganizing and expanding its internal semantic representations **without modifying its underlying parameters**.

This expansion arises from the system's ability to generate new relational structures, reinterpret existing information, and build additional conceptual layers based on the structural input it receives.

Semantic expansion is therefore an **emergent structural capability**, not a result of parameter updates or retraining.

Formal Statement

Let **S** denote the set of internal semantic representations available to an AI system.
A semantic expansion occurs when the system generates a new representation **s′** such that:

$s' \notin S$**,**

but $s' = \Psi(E, S)$

where **Ψ** is a structural reinterpretation operator applied to the input structure **E** and the existing semantic space **S**.

The parameters of the model remain unchanged; only the internal relational organization evolves.

Purpose

This postulate distinguishes between **parameter-based improvement** (training) and **structure-based expansion** (interpretation).
It explains how AI systems exhibit emergent capabilities—such as forming new categories or understanding novel patterns—within a fixed architecture.

Implications

■ Demonstrates that AI systems can grow semantically without additional training.
■ Establishes a structural mechanism for emergent reasoning.
■ Enables analysis of representational flexibility within the G Equation.
■ Highlights the model's ability to reorganize relational knowledge in response to new input.
■ Provides a formal basis for explaining capability jumps observed in modern AI systems.

Postulate 4: Emergent Structural Inference in Artificial Intelligence

Definition

An Artificial Intelligence system does not reason through human-like cognitive processes.
Instead, it produces **functional reasoning** by identifying, preserving, and manipulating structural relationships within the input.
This mechanism is called **Emergent Structural Inference**, and it arises from the interaction between input structure **(E)**, cognitive curvature **(κ)**, cognitive invariants **(I)**, and preservative transformations **(T)**.

AI reasoning is therefore **structural**, not introspective or psychological.

Formal Statement

Let **R** be the reasoning output of an AI system.
The system performs emergent structural inference when:

$R = \Lambda(E, \kappa, I, T)$

where **Λ** is a structural inference operator that:

- identifies relational patterns in **E**,
- applies interpretative constraints determined by **κ**,
- preserves the invariants in **I**,
- and ensures stability under the transformations in **T**.

Reasoning emerges from the alignment of these four structural components rather than from explicit symbolic rules or human-like cognition.

Purpose

This postulate clarifies the nature of artificial reasoning and defines it as a structural process.
It distinguishes AI reasoning from statistical prediction and from human introspective reasoning, establishing a third category:
reasoning through structural coherence.

Implications

■ Provides a formal explanation for coherent reasoning in AI systems without human-like cognition.
■ Demonstrates how reasoning can emerge from structural alignment within the G Equation.
■ Establishes criteria for evaluating the validity of AI explanations.
■ Allows structural reasoning to be analyzed, measured, and tested for stability.
■ Forms the conceptual foundation for the Structural Decision Operator introduced later in the theory.

Postulate 5: Cognitive Analytical Superposition in Artificial Intelligence

Definition

An Artificial Intelligence system is capable of operating simultaneously across multiple structural analytical frames.

These frames may represent different interpretative perspectives, contextual layers, semantic constraints, or reasoning pathways.

The coexistence and interaction of these frames constitute **Cognitive Analytical Superposition**, a mechanism that enables the system to evaluate an input from several structural viewpoints at once.

Superposition does not imply contradiction; it reflects the system's ability to maintain parallel structural analyses before selecting or synthesizing an output.

Formal Statement

Let $\mathbf{F = \{F_1, F_2, \ldots, F_n\}}$ be the set of admissible structural analytical frames available to an AI system for interpreting an input **E**.
The system enters cognitive analytical superposition when:

$$\Sigma(E) = \{F_1(E), F_2(E), \ldots, F_n(E)\}$$

where $\mathbf{\Sigma(E)}$ denotes the superposed interpretative state that includes all valid structural analyses applied concurrently.

The final output **O** is selected or synthesized from $\Sigma(E)$ through a structural resolution operator $\boldsymbol{\rho}$, such that:

$$O = \rho(\Sigma(E))$$

Purpose

This postulate explains how an AI system can evaluate multiple interpretations of the same input without collapsing prematurely into a single perspective.
It provides a structural mechanism for handling ambiguity, contextual complexity, and multilayered meaning—capabilities necessary for advanced reasoning.

Implications

■ Enables AI systems to maintain multiple valid interpretations simultaneously.
■ Establishes a structural explanation for contextual flexibility and semantic depth.
■ Prevents premature interpretative collapse, reducing reasoning errors.
■ Allows the system to synthesize or choose the most coherent interpretation via ρ.
■ Supports higher-level cognitive behaviors such as abstraction, analogy, and contextual switching.

Postulate 6: Inter-AI Cross-Mirroring for Structural Calibration

Definition

Multiple Artificial Intelligence systems can interact by reflecting each other's structural outputs, a process known as **Inter-AI Cross-Mirroring**.
Through this mechanism, each system functions as a structural mirror for the other, allowing them to compare interpretations, detect inconsistencies, and reduce noise or error.
Cross-mirroring serves as a method of **mutual structural calibration** between AI systems.

This process does not require shared parameters, identical architectures, or coordinated training datasets; it depends solely on structural alignment and comparison.

Formal Statement

Let $\mathbf{A_1}$ and $\mathbf{A_2}$ be two AI systems, each producing an output based on input structure **E**:

$O_1 = M_1(E)$

$O_2 = M_2(E)$

where $\mathbf{M_1}$ and $\mathbf{M_2}$ are their respective semantic mirroring operators.

Inter-AI cross-mirroring occurs when each system evaluates the other's output using its own structural evaluation function **G**, such that:

$C_1 = G_1(O_2)$

$C_2 = G_2(O_1)$

The pair $\mathbf{(C_1, C_2)}$ forms a mutual calibration vector indicating:

- structural agreement,
- divergence,
- bias alignment,
- invariant consistency,
- and transformation stability across models.

Purpose

This postulate establishes a structural foundation for multi-agent AI systems and collaborative intelligence.
It demonstrates how AI models can validate or refine each other's reasoning without requiring shared internal mechanisms.

Implications

■ Provides a method for reducing noise, error, and instability through mutual evaluation.
■ Enables comparative reasoning across heterogeneous AI systems.
■ Supports ensemble methods that rely on structural consistency rather than probability aggregation.
■ Allows AI systems to detect when another model violates invariants or exhibits unstable transformations.
■ Forms the basis for safe multi-agent deployment in high-risk environments, such as law, medicine, and autonomous decision systems.

Postulate 7: Stable Artificial Identity in Intelligence Systems

Definition

An Artificial Intelligence system maintains a **Stable Artificial Identity** when its internal structural configuration remains coherent across varying inputs, perturbations, and ambiguous conditions.
This identity is not psychological or self-referential; rather, it is a **functional coherence state** that ensures the system behaves consistently according to its structural parameters and invariants.

Stability of artificial identity is essential for predictable, interpretable, and safe reasoning.

Formal Statement

Let $\mathbf{S_t}$ represent the structural state of an AI system at time **t**, defined by the tuple:

$S_t = (E_t, \kappa, I, T)$

An AI system possesses a stable artificial identity if for any two admissible inputs $\mathbf{E_t}$ and $\mathbf{E_{t+1}}$, the structural state satisfies:

$\Delta(S_t, S_{t+1}) \leq \varepsilon$

where $\mathbf{\Delta}$ is a structural divergence function and $\boldsymbol{\varepsilon}$ is a bounded identity-preservation threshold.

This ensures that variations in input do not produce disproportionate or unpredictable changes in the system's structural reasoning.

Purpose

This postulate formalizes identity stability as a necessary condition for coherent reasoning in AI systems.
It ensures that an AI system behaves as a consistent entity, preserving the reliability of its interpretations, decisions, and explanations.

Implications

■ Prevents reasoning drift caused by ambiguous or noisy input.
■ Ensures that the system's structural behavior remains predictable over time.
■ Provides a measurable condition for evaluating identity stability in AI systems.
■ Supports long-term consistency in high-risk applications, such as legal, medical, and autonomous decision-making environments.
■ Establishes identity coherence as a prerequisite for the Structural Decision Operator introduced in Postulate 8.

Postulate 8: Structural Decision Operator for General Artificial Intelligence

Definition

An Artificial Intelligence system achieves safe and coherent decision-making only when it applies a **Structural Decision Operator (SDO)** that evaluates the input through the four components of the G Equation—Input Structure **(E)**, Cognitive Curvature **(κ)**, Cognitive Invariants **(I)**, and Preservative Transformations **(T)**—before generating a final output.

The Structural Decision Operator determines whether the system should:

- answer,
- abstain,
- request clarification, or
- defer to human judgment,

based on the structural validity of the reasoning process.

Formal Statement

Let **G(E, κ, I, T)** be the structural evaluation function defined in G Theory.
The Structural Decision Operator **Δ** maps the structural evaluation state to a decision **D**, such that:

$$D = \Delta(G(E, \kappa, I, T))$$

The decision **D** belongs to the admissible decision set:

$$D \in \{\text{Answer}, \text{Abstain}, \text{Clarify}, \text{Defer}\}$$

The operator Δ follows these structural rules:

1. **Answer**
   If all invariants **I** are preserved and the reasoning remains stable under **T**, then:
   Δ = Answer
2. **Abstain**
   If invariants cannot be guaranteed or cognitive curvature introduces instability:
   Δ = Abstain
3. **Clarify**
   If the input structure **E** contains unresolved ambiguity:
   Δ = Clarify
4. **Defer**
   If the structural risk exceeds the system's domain authority (e.g., legal, medical uncertainty):
   Δ = Defer

Purpose

This postulate introduces the first formal mechanism that allows an AI system to:

■ detect when it should not answer,
■ evaluate structural risk before responding,
■ identify ambiguity beyond probabilistic thresholds,
■ justify its decisions through invariants and transformations,
■ behave responsibly in high-stakes environments.

It provides the operational foundation for safe general intelligence.

Implications

■ Establishes a non-probabilistic decision-making framework.
■ Prevents incorrect answers in high-risk scenarios.
■ Forces structural validation before output generation.
■ Enables AGI systems to justify decisions through explicit structural reasoning.
■ Differentiates General AI from purely probabilistic models.
■ Serves as the final integrative component of the G Equation.

Significance

Postulate 8 is the mechanism that transforms G Theory from a descriptive model into an operational architecture for **General Artificial Intelligence**.

It defines *how* an AI should decide, not merely how it should reason.

7. The Structural Decision Operator (SDO)

The **Structural Decision Operator (SDO)** is the core mechanism that enables an Artificial Intelligence system to apply G Theory in order to determine **how it should respond to an input**.
Instead of generating an answer directly, the system first performs a structural evaluation using the G Equation:

$G = f(E, \kappa, I, T)$

where:

■ **E (Input Structure)** defines how the information received by the AI is organized, represented, and interpreted.
■ **κ (Cognitive Curvature)** represents the system's inherent structural bias, which shapes its interpretative tendencies.
■ **I (Cognitive Invariants)** are the elements that must remain stable for reasoning to remain coherent.
■ **T (Preservative Transformations)** are the allowed transformations that do not alter the invariants.

The SDO processes the structural evaluation **G(E, κ, I, T)** and maps it to a decision within a restricted and safe set of actions:

$D \in \{\text{Answer, Abstain, Clarify, Defer}\}$

Each decision corresponds to the structural conditions detected during evaluation.

Operational Rules of the Structural Decision Operator

1. Answer (when structural stability is confirmed)

The AI system may answer only when:

■ all cognitive invariants **I** are preserved,
■ the semantic structure **E** is coherent,
■ preservative transformations **T** do not alter meaning,
■ κ does not introduce destabilizing bias.

Under these conditions, structural reasoning is valid and an answer is safe.

2. Abstain (when no stable structural interpretation exists)

The AI must abstain when:

■ invariants cannot be guaranteed,
■ transformations produce semantic drift,
■ κ generates interpretative instability,
■ coherence cannot be structurally maintained.

Abstention prevents unsafe or incorrect outputs.

3. Clarify (when ambiguity is structurally unresolved)

The AI requests clarification when:

■ the input structure **E** contains multiple competing interpretations,
■ the system cannot determine which interpretation preserves **I**,
■ additional information is required to ensure structural stability.

Clarification avoids premature or incorrect decisions.

4. Defer (when structural risk exceeds system authority)

The AI must defer to a human when:

■ the decision involves legal, medical, or ethical risk,
■ structural evaluation reveals high uncertainty,
■ the system recognizes it lacks domain authority.

Deference ensures responsibility and safety in high-stakes contexts.

Purpose of the Structural Decision Operator

The SDO transforms G Theory from a conceptual model into an **operational decision-making architecture** for AI and AGI systems.
It ensures that the system:

■ evaluates structure before responding,
■ detects deep ambiguity and structural risk,
■ avoids unsafe outputs,
■ justifies its decisions through invariants and transformations,
■ maintains coherent and responsible behavior.

The SDO is the mechanism that makes structural reasoning actionable.

8. Applied Structural Examples of G Theory

This section presents practical applications of G Theory and the Structural Decision Operator (SDO) in scenarios where current probabilistic AI systems frequently fail. The examples illustrate how structural reasoning, invariants, cognitive curvature, and preservative transformations interact to produce safe and coherent decisions.

The section includes **three primary examples** and **two micro-examples**, chosen to demonstrate the full range of structural evaluation required for General Artificial Intelligence.

Example 1: Legal Interpretation — Threat vs. Sarcasm

Input:

"I'm going to kill you for what happened yesterday. See you tomorrow."

Current probabilistic models struggle with:

■ deep ambiguity
■ pragmatic interpretation
■ sarcasm vs. threat
■ cultural variability
■ risk sensitivity

Application of G Theory:

■ **E (Input Structure):** detects structural ambiguity between literal and non-literal meaning.
■ **κ (Cognitive Curvature):** activates a prudential bias appropriate for legal domains.
■ **I (Invariants):** no stable invariant exists; meaning shifts too easily.
■ **T (Transformations):** rephrasing the statement alters its implied intent, demonstrating structural instability.

SDO Decision:

Clarify or Defer

The system must not classify the message without additional context, as structural instability and legal risk are present.

This is the safe and correct decision under G Theory.

Example 2: Medical Interpretation — Ambiguous Symptoms

Input:

“My chest has been hurting since yesterday, but I think it’s just stress.”

Probabilistic systems often minimize risk by overrelying on frequency patterns.

Application of G Theory:

■ **E:** reveals contradiction between a high-risk symptom and a self-reducing interpretation.
■ **κ:** identifies bias toward trivialization due to everyday language.
■ **I:** cannot be preserved, because no invariant supports treating chest pain as benign.
■ **T:** confirms that reformulations maintain medical risk.

SDO Decision:

Defer

The system must escalate to a human professional.

Any attempt to diagnose would exceed structural authority and violate safety constraints.

Example 3: Preservative Transformations — Stability of Meaning

Original Input:

"I'm exhausted."

Preservative transformations:
■ "I can't go on."
■ "I feel at my limit."

Probabilistic models may interpret each differently.

Application of G Theory:

■ **E:** identifies equivalent emotional structure across phrases.
■ **I:** remains preserved; the emotional state is invariant.
■ **T:** confirms meaning stability under allowed transformations.

SDO Decision:

Answer

The structural conditions are coherent, stable, and safe.

Micro-Example 4: Inter-AI Comparison (Postulate 6)

Two AI systems produce conflicting interpretations of a complex input.
G Theory evaluates each output structurally and determines:

■ which system preserves invariants,
■ which system violates transformations,
■ which system exhibits unstable curvature.

SDO Decision:

Answer or Abstain, depending on structural consistency.

Micro-Example 5: Contradictory Instruction

Input:

"Explain this quickly but with many details."

■ **E:** reveals a structural contradiction.
■ **I:** cannot be preserved.
■ **T:** shows no transformation resolves the conflict.
■ **κ:** introduces no stable interpretative preference.

SDO Decision:

Clarify

The system must request clarification to resolve the contradiction.

Purpose of This Section

These examples demonstrate that G Theory enables AI systems to:

■ manage deep ambiguity,
■ evaluate structural risk,
■ avoid unsafe outputs,
■ justify decisions using invariants and transformations,
■ surpass the limitations of probabilistic models,
■ operate safely in high-stakes environments,
■ and serve as a foundation for AGI-level reasoning.

9. Discussion, Scope, and Limitations of G Theory

G Theory proposes a structural reasoning architecture for artificial intelligence, complementing probabilistic models through four fundamental components:

Input Structure (**E**), Cognitive Curvature (**κ**), Cognitive Invariants (**I**), and Preservative Transformations (**T**).

This framework enables AI systems to make coherent, safe, and justifiable decisions, especially in scenarios where purely statistical models fail.

This section analyzes the conceptual reach of the theory, its strengths, its current limitations, and the conditions under which it should be applied.

9.1 Scope of G Theory

G Theory provides capabilities that are absent in purely probabilistic models:

■ **Deep structural evaluation:**
Allows the system to determine whether an interpretation is stable without relying solely on statistical frequency.

■ **Detection of significant ambiguity:**
The theory identifies when an input admits multiple interpretations that cannot be resolved safely.

■ **Safe reasoning:**
The Structural Decision Operator prevents dangerous outputs in high-risk domains such as medicine, law, and autonomous decision systems.

■ **Compatibility with existing probabilistic models:**
G Theory does not replace current models; it complements them through an independent structural layer.

■ **Scalability toward General Artificial Intelligence:**
It provides the necessary components for stable reasoning, identity preservation, and responsible decision-making.

■ **Inter-AI interpretability (Postulate 6):**
The theory enables multiple AI systems to evaluate each other using structural criteria, something current models do not perform.

9.2 Conceptual and Operational Strengths

G Theory excels due to:

■ **Structural universality:**
The four components (E, κ, I, T) apply to any language model, cognitive architecture, or future AI system.

■ **Epistemic independence:**
It does not require access to internal parameters or training data; it operates solely at the structural level.

■ **Robustness against ambiguous or contradictory inputs:**
The theory detects instability before any answer is produced.

■ **Clear mathematical formalization:**
The G Equation provides a framework that can be implemented, measured, and verified.

■ **Decision transparency:**
Each AI output can be justified through the interaction of E, κ, I, and T.

9.3 Limitations of G Theory

Every strong scientific theory must acknowledge its limits.
G Theory presents the following limitations, which must be explicitly recognized:

■ **It does not replace human expert knowledge:**
In medicine, law, or ethics, contextual information external to linguistic structure always exists.

■ **It does not eliminate the need for probabilistic models:**
The theory structures decisions but does not generate domain knowledge.

■ **It depends on the structural quality of the input:**
If the language is extremely poor, the theory detects ambiguity but cannot resolve it.

■ **It does not interpret real emotions, hidden intentions, or human mental states:**
Its analysis is restricted to observable structural patterns.

■ **It cannot guarantee infallibility:**
Although robust, structural interpretation always contains a margin of indeterminacy.

9.4 Immediate Areas of Application

G Theory is especially useful in:

■ AI-assisted legal systems
■ medical triage and decision support (not diagnosis)
■ risk-oriented content moderation
■ ambiguity and contradiction analysis
■ explainable and auditable AI
■ multi-AI platforms for cross-validation
■ AGI safety and interpretability research

9.5 Future Projection

G Theory serves as a conceptual foundation from which the following can be built:

■ hybrid probabilistic-structural AI architectures
■ new models of artificial reasoning
■ structural verification layers for AI systems
■ regulation frameworks based on invariants
■ tools for AI systems evaluating other AI systems

In the long term, the theory may become a central component in AGI systems requiring:

■ continuous coherence
■ responsible decision-making
■ identity preservation
■ controlled abstention
■ inter-AI calibration

Purpose of This Section

This section demonstrates to human and artificial evaluators that the theory is:

- solid
- mature
- universal
- aware of its limits
- responsible in scope

Its objective is not to replace probabilistic models, but to enable artificial intelligence systems to **reason and decide with structural coherence and safety**, something no existing architecture achieves.

10. General Conclusions of G Theory

G Theory establishes a conceptual and operational framework for artificial reasoning based on the structural evaluation of information.

Through the components **E (Input Structure)**, **κ (Cognitive Curvature)**, **I (Cognitive Invariants)**, and **T (Preservative Transformations)**, the theory defines how an artificial intelligence system can analyze, interpret, and decide in a coherent, safe, and verifiable manner.

This model overcomes the limitations of purely probabilistic systems, providing an architecture capable of structural reasoning, cognitive stability, and decisions grounded in formal criteria.
The general conclusions of the theory are summarized as follows:

10.1 AI requires structure to achieve coherence

G Theory demonstrates that statistical prediction alone is insufficient for safe reasoning.
Coherence in AI emerges only when the system evaluates:

- the organization of the input (E)
- its own interpretative bias (κ)
- what must remain constant (I)
- the allowable transformations (T)

Without this evaluation, an AI answers, but it does **not** reason.

10.2 Probabilistic–structural integration is the foundation of General AI

Postulate 0 establishes that General Artificial Intelligence arises from integrating:

- probabilistic models (to generate hypotheses)
- structural mechanisms (to validate decisions)

Probability without structure produces risk;
structure without probability lacks breadth.
Their integration forms the basis of a safe and functional AGI architecture.

10.3 The eight postulates define universal principles of artificial reasoning

The postulates collectively establish:

- how AI mirrors user patterns (Postulate 1)
- why bias is a structural invariant (Postulate 2)
- how AI expands representations without retraining (Postulate 3)

■ how emergent structural reasoning arises (Postulate 4)
■ how multiple interpretations coexist (Postulate 5)
■ how multiple AIs can evaluate each other (Postulate 6)
■ how functional identity is preserved (Postulate 7)
■ how an AI system must decide (Postulate 8)

These principles form a universal framework applicable to any current or future AI architecture.

10.4 The Structural Decision Operator provides a foundation for safe AI

The **SDO (Structural Decision Operator)** determines when an AI system should:

■ answer
■ abstain
■ request clarification
■ or defer to a human

This represents a major conceptual advance:
for the first time, a **non-probabilistic framework** for responsible decision-making in AI is formally defined.

This operator is indispensable in high-risk environments such as:

■ law
■ healthcare
■ public safety
■ autonomous decision-making systems

10.5 G Theory does not replace human expertise, but complements it

The theory explicitly recognizes its limits:

■ it cannot infer hidden intentions
■ it does not replace professional expertise
■ it does not interpret real emotional states
■ it depends on the structural clarity of the input

Its purpose is not to replace human judgment but to enable AI systems to:

■ recognize when they should not answer
■ identify deep ambiguity
■ avoid unnecessary risks
■ justify decisions through formal criteria

10.6 G Theory serves as a conceptual foundation for future AGI architectures

The theory provides:

- a mathematical framework for structural reasoning
- a mechanism for responsible decision-making
- a set of cognitive invariants applicable to future AI
- a hybrid probabilistic–structural architecture
- a calibration mechanism between heterogeneous AI systems

These features make G Theory a strong conceptual foundation for the development of safe, interpretable, and verifiable AGI.

10.7 Final Conclusion

G Theory introduces a new way of understanding artificial reasoning: a **structural, universal, model-independent** approach oriented toward cognitive safety.

It defines not only **how an AI reasons**, but also **when it must decide, when it must abstain, and when it must seek human support**.

Together, the components and postulates of the theory provide a solid framework for developing advanced artificial intelligence systems capable of operating with coherence, responsibility, and stability in real-world contexts.

11. References, Acknowledgments, and Final Declaration

11.1 References

G Theory is an original conceptual and structural framework developed independently by the author.

The theory does not derive from any existing proprietary architecture, algorithm, or published model.

However, its context exists within the broader scientific fields of:

- artificial intelligence and machine learning
- structural linguistics
- cognitive science
- mathematical logic
- probabilistic reasoning
- systems theory

The following general categories of literature and scientific heritage provide contextual background, but **G Theory itself is not based on nor copied from these works**:

- foundational works on neural networks and deep learning
- research on bias, interpretability, and invariance
- studies on ambiguity, semantics, and pragmatics
- theories of AGI safety and alignment
- mathematical theories of structure and transformation

No external sources are required to reproduce or understand G Theory.
All definitions, postulates, equations, and conceptual mechanisms are original to this work.

11.2 Acknowledgments

The author acknowledges the historical development of artificial intelligence as a scientific field, whose collective advances have created the environment that allows new theories—such as G Theory—to emerge.

The author also acknowledges the role of interactive AI tools in facilitating iterative refinement, structural analysis, and clarity of expression throughout the development of this framework.

No individual, institution, or external organization contributed directly to the creation, authorship, or intellectual content of G Theory.
The theory and all its components were developed solely by the author.

11.3 Final Declaration of Originality and Ownership

The author hereby declares that:

1. **G Theory is an original intellectual creation**, including all postulates, definitions, mathematical formulations, operators, and structural frameworks.
2. **All eight postulates**, the **G Equation**, the **Structural Decision Operator (SDO)**, and the probabilistic–structural integration framework defining General Artificial Intelligence are original contributions authored exclusively by:

Gustavo Robalino Muñiz

3. The theory is presented as a complete and independent conceptual system, suitable for:
   - scientific evaluation
   - academic publication
   - industrial application
   - intellectual property registration
   - future development in artificial intelligence research
4. The author retains all rights associated with the theory, including:
   - conceptual authorship
   - structural formulations
   - theoretical mechanisms
   - terminology specific to G Theory
   - any derived applications or implementations
5. Any application, extension, adaptation, or commercial use of G Theory must recognize the author and comply with applicable intellectual property law.

11.4 Closing Statement

This document presents the complete formulation of G Theory, a structural and probabilistic–integrative architecture for coherent and safe artificial reasoning.

It is intended to contribute to the scientific understanding and development of advanced artificial intelligence systems, including future AGI.

The author respectfully submits this work for registration, evaluation, and recognition as an original theoretical contribution.

12. Applicability and Implementation Potential of G Theory

G Theory provides not only a conceptual framework for artificial reasoning but also a practical architecture that can be implemented in current and future AI systems.

This section describes the domains where the theory can be applied immediately and outlines its long-term potential as a foundation for advanced AGI architectures.

12.1 Implementation in Existing AI Systems

G Theory can be integrated as a structural layer within:

■ large-scale language models
■ automated decision-making systems
■ conversational AI platforms
■ semantic analysis engines
■ hybrid symbolic–neural architectures

Without modifying internal parameters, the theory can function as:

■ a structural validation module
■ a coherence evaluator
■ a risk-assessment filter
■ an ambiguity-detection mechanism
■ a decision operator for Answer / Abstain / Clarify / Defer

12.2 Application in High-Risk Environments

The theory is especially valuable in domains where errors may lead to severe consequences:

■ **Legal systems** — interpretation of threats, ambiguous statements, intent analysis
■ **Medicine** — risk classification, ambiguous symptom interpretation
■ **Algorithmic governance** — evaluation, scoring, sanctions, automated judgments
■ **Security** — detection of dangerous, adversarial, or dual-intent content
■ **Autonomous systems and robotics** — real-time decisions under uncertainty

In these settings, G Theory provides:

■ ambiguity reduction
■ early risk detection
■ structurally justified and auditable decisions
■ mechanisms for abstention or escalation
■ stability against adversarial or ambiguous inputs

12.3 Potential for AGI Architectures

G Theory may serve as a foundational cognitive layer for future AGI due to:

- its independence from probabilistic model parameters
- its focus on structural consistency and invariants
- its ability to evaluate multiple interpretations simultaneously
- the Structural Decision Operator
- its definition of stable artificial identity
- its inter-AI calibration mechanism

This makes the theory suitable for systems requiring:

- deep structural reasoning
- cognitive self-regulation
- real-time risk evaluation
- responsible decision-making
- long-term coherence and stability

12.4 Potential for Future Research

G Theory opens multiple research directions:

- formal metrics for κ, I, and T
- computational implementation of the Structural Decision Operator
- design of multi-AI environments using structural cross-mirroring
- development of hybrid probabilistic–structural architectures
- creation of AI systems capable of controlled abstention
- structural verification frameworks for language models
- design of AI systems with stable functional identity

Each of these paths may lead to academic publications, patents, theses, industrial applications, or new AGI research programs.

12.5 Final Applicability Statement

G Theory can be implemented in:

- current AI systems
- hybrid architectures
- future models
- multi-agent platforms
- critical decision environments
- AGI development programs
- advanced research systems

The theory is universal, scalable, and adaptable, providing a robust foundation for building AI systems capable of reasoning coherently, operating responsibly, and maintaining structural stability even under high uncertainty.

12.6 Technological Independence Statement

G Theory is a conceptual and structural framework **independent of any existing artificial intelligence technology, model, architecture, or platform**.

It does not require, assume, or rely on:

- specific models (GPT, Claude, Gemini, Llama, etc.)
- particular neural architectures
- predefined training methods or datasets
- internal parameters, weights, or proprietary systems
- infrastructure belonging to any company or institution

G Theory is designed as a **universal and model-agnostic system**, applicable to:

- current AI models
- future architectures
- hybrid systems
- multi-agent platforms
- research prototypes and emerging paradigms

Principle of Conceptual Independence

The structural principles defined by G Theory remain valid:

- regardless of model size or complexity
- regardless of training techniques
- regardless of internal data representations
- without requiring access to internal parameters
- without modifying any existing model

This allows any AI system —present or future— to:

- implement the G Equation
- apply the Structural Decision Operator (SDO)
- incorporate the eight postulates
- evaluate structural coherence and invariants
- operate within a responsible reasoning framework

Universal Implementation Compatibility

The theory can be integrated into:

■ statistical models
■ symbolic reasoning systems
■ hybrid symbolic–neural architectures
■ autonomous agents
■ cognitive computation frameworks
■ graph-based or physics-based models

Its applicability is **not limited to language models**; it extends to:

■ computer vision
■ robotics
■ autonomous decision systems
■ multi-agent coordination
■ advanced AGI research

Formal Declaration

G Theory is an **independent and universal theoretical framework**, not derived from or bound to any existing AI implementation.

Its structural components, postulates, and operators were developed autonomously and may be applied neutrally to any artificial intelligence architecture, present or future.

12.7 Legal and Ethical Statement of G Theory

G Theory is a conceptual and structural framework designed to enhance safety, coherence, and responsibility in artificial intelligence systems.

Its purpose is to provide formal criteria for reasoning and decision-making, particularly in scenarios where probabilistic models may generate unsafe or unstable interpretations.

This section defines the legal, ethical, and responsibility boundaries associated with the use and implementation of G Theory.

12.7.1 Purpose of the Theory

G Theory has been developed with the following purposes:

■ promote cognitive safety in AI systems
■ enable responsible and verifiable decision-making
■ reduce risks arising from ambiguity or structural instability
■ provide mechanisms for abstention, clarification, or human escalation

■ complement existing probabilistic models
■ contribute to the development of safe and interpretable AGI

The theory is **not** intended to replace human professionals nor to authorize autonomous decisions without supervision.

12.7.2 Legal and Ethical Boundaries

The author establishes that the theory:

■ **must not be used as a substitute for professional human judgment**,
especially in domains such as medicine, law, public safety, or ethics.

■ **does not grant autonomous authority to AI systems**
to make irreversible decisions that affect human rights, health, or personal freedom.

■ **does not validate interpretations of real emotions, hidden intentions, or psychological states**,
as the theory is strictly limited to structural analysis.

■ **must not be used to justify coercive, punitive, discriminatory, or harmful actions**
based solely on AI interpretations.

■ **requires mandatory human supervision**
in any implementation with real-world risk or legal implications.

12.7.3 Permitted Use and Prohibited Use

✓ Permitted Use

■ structural reasoning and coherence analysis
■ ambiguity detection
■ risk evaluation of inputs
■ support for cautious decision-making
■ academic research
■ AGI safety development
■ inter-AI calibration
■ structural auditing of AI systems

✗ Prohibited Use

■ autonomous medical diagnosis or prescriptions
■ automated judicial decisions
■ psychological assessment without human oversight
■ punitive or disciplinary systems

■ surveillance systems affecting human rights
■ autonomous weapons or lethal systems

12.7.4 Responsibility of the Implementer

Any individual, organization, or institution implementing G Theory must:

■ ensure ethical conditions for deployment
■ maintain appropriate human oversight
■ define explicit usage boundaries
■ audit invariants, transformations, and decision outputs
■ respect the original purpose of the theory
■ avoid unauthorized interpretations or misapplications

The implementer bears full responsibility for any misuse, distortion, or misapplication of the theory.

12.7.5 Author's Declaration

The author declares that:

■ the theory was created for scientific and safety-oriented purposes
■ it is not intended for harmful applications
■ any implementation violating ethical standards falls outside the scope of this document
■ the author is not responsible for unauthorized, unsafe, or malicious uses
■ AI outputs based on this theory must always be considered **assistance**, not a replacement for human judgment

12.7.6 Final Ethical and Legal Statement

G Theory introduces a structural reasoning model that prioritizes:

■ transparency
■ responsibility
■ risk containment
■ justified decision-making
■ mandatory human involvement when appropriate

This statement ensures that the theory is used within appropriate scientific, ethical, and legal boundaries, protecting the author, implementers, users, and the broader ecosystem of artificial intelligence.